Peer Disagreement in Cosmology

Formal Methods II
Sebastián Gil Rodríguez*

November 2, 2019

This paper presents a robustness analysis of the agent-based model (hereafter ABM) developed by Douven 2010 and applies it to the field of physical cosmology. The simulations discussed below are meant to realistically model the current discrepancy in the measured values of the Hubble constant as of July 2019. The implementation of the ABM is divided into an observation stage and a conference stage. The paper is organized as follows. I first give an overview of the philosophy of peer disagreement and summarize the main results of Douven's paper. Next, I look at the technical aspects of Douven's ABM and motivate its application to the parameter space of Hubble constant measurements. I then explain my expectations for the simulations and discuss whether they were met by the results obtained.

1 Simulating Peer Disagreement

In the social epistemology literature, *peer disagreement* refers to situations where epistemic agents disagree on a given issue even though they all have access to the same evidence and share the same expertise in the field of contention. The basic question is whether it is preferable for epistemic peers to adopt an intermediate belief or to refuse to change their minds. Some philosophers argue that individuals should always *split the difference*: they ought to assign equal weights to the beliefs of their peers and average them to arrive at a new one. Identifying the *Irrationality Claim* (IC) as the claim that rationality requires epistemic peers to always split the difference, Douven begins his paper by pointing out that IC's proponents seem to undermine the view that "science is a paradigmatically rational enterprise".[1] After all, the history of science is full of examples of profound disagreements among scientists unwavering in their commitment to their chosen views. The fact that science has nonetheless been successful is the ground for Douven's criticism of IC's validity.

Douven aims to undermine the presupposition that the validity of IC can be a priori established by presenting computer simulations in which the rational response to peer disagreement is dependent on highly context-sensitive and contingent factors. The simulations are based on an extension of the Hegselmann-Krause model of opinion dynamics consisting of agents trying to find the correct value of a physical parameter in the presence of noisy data. In particular, he

*Munich Center for Mathematical Philosophy, Matrikelnummer: 12011085
[1] Douven 2010, p149.

compares agents that split the difference with their peers to agents that let themselves be guided only by the data. The main result is that no difference-splitting communities approach the correct parameter value very rapidly without further convergence in their beliefs. Conversely, difference-splitting communities converge significantly closer to the true parameter value but take much longer to do so.

To interpret his results, Douven defines epistemic rationality in terms of conduciveness to the *epistemic goal* of wanting "to believe the truth and nothing but the truth."[2] Given this goal, it would seem that both difference-splitting and no difference-splitting can be rational responses to peer disagreement depending on whether agents require convergence to the truth with very high precision or prefer a less precise but much faster convergence. Such a choice is based on contingent factors, which is sufficient to meet Douven's goal of demonstrating that the validity of IC cannot be established *a priori*.

A proponent of IC may still object that Douven's conclusions are highly sensitive to the type of simulation employed and that the simulations may yield different results upon modification of their ideal assumptions and parameters. Anticipating this objection, Douven retorts that "the simulations presented in this paper are not meant to model reality to any degree of precision," but merely to model situations that "certainly *could* occur."[3] In light of this, my aim here is to use Douven's simulations to precisely model a case of peer disagreement that has already occurred—the Hubble tension in physical cosmology—and to explore how the disagreement may be resolved in a scientific conference. But first, we'll look at the Hegselmann-Krause ABM in more detail.

2 The Hegselmann-Krause Model

Consider a set of n agents individually trying to determine the true value of some parameter τ. Letting $x_i(u)$ denote the i-th agent's belief of τ after the u-th update, the Hegselmann-Krause model updates agent beliefs with the following rule:

$$x_i(u+1) = \alpha \frac{1}{|X_i(u)|} \sum_{j \in X_i(u)} x_j(u) + (1-\alpha)\tau \tag{1}$$

The rule sums over the belief of all agents j belonging to the following set:

$$X_i(u) := \{j : |x_i(u) - x_j(u)| \leq \varepsilon\} \tag{2}$$

Recognizing $|X_i(u)|$ as the cardinality of the set, we see that the first term in (1) is a weighs by a factor α the average of all agents in (2). The ε parameter in (2) defines the size of the tolerance interval over which agents are willing to "listen" to their peers: the larger its value, the more agents are included in the weighted average. Now, the second term in the rule represents the agent's "measurement" of the parameter τ such that for all non-zero values of α and ε, the agent will take into account both the opinion of the neighbors in its tolerance interval and its own individual measurements.

[2] Douven infers this goal from (Rescher 1973), p. 21, (Lehrer 1974), p. 202, (BonJour 1985), p. 8, and (Foley 1992), p. 183. See the references in (Douven 2010)

[3] Douven 2010, p.155.

Douven's main modification to the above rule consists of adding a uniformly-distributed random variable ζ to the second term, resulting in the expression

$$x_i(u+1) = \alpha \frac{1}{|X_i(u)|} \sum_{j \in X_i(u)} x_j(u) + (1-\alpha)(\tau + \mathrm{rnd}(\zeta)) \tag{3}$$

The idea here is that, instead of directly interacting with the true value of τ, agents receive a noisy signal from $\mathrm{rnd}(\zeta)$ every time they take a measurement. Table 1 summarizes the parameters held constant in Douven's simulations:

Parameter	Explanation	Value Assigned
n	Number of agents	25
u	Number of updates	50
τ	True parameter value	0.75
ζ	Noise level	0.2

Table 1: Parameters held fixed in Douven's simulations

As described in the previous section, the core of Douven's analysis consists of comparing difference-splitting communities to no difference-splitting communities. The parameter space for both scenarios is summarized in Table 2.

Parameter	Explanation	No difference-splitting	Difference-splitting
ε	Size of tolerance interval	0.0	0.1
α	Weight given to peers' opinion	0.5	0.9

Table 2: Parameters used to assess the validity of *IC*

With $\varepsilon = 0$, agents will only update their beliefs in terms of individual measurements, so the above parameter choices correctly distinguish difference-splitting scenarios from no difference-splitting ones. The parameters in Table 1, however, have been arbitrarily chosen and one is left to wonder why τ, ζ and ε—which should model the physical constant in question—have all been drawn from the real number interval $[0, 1]$. Because these parameters are meant to model a physical quantity, they should all have the dimensions of whatever is being measured. For example, if τ is chosen to represent the value of the Hubble constant, the theory and history of which are discussed in Appendix A, the other parameters n, ζ, and ϵ should be suitably modified to model the peculiarities of the cosmology community. This is the subject of the next section.

3 My ABM

The main challenge towards developing a realistic ABM simulation of the Hubble tension is recognizing that the Hegselmann-Krause model presupposes the availability of one "true" value of τ which, while not intended to be known by the agents *a priori*, drives the evolution of their opinion dynamics towards a unique end-state. Some cosmologists believe that the resolution of the Hubble tension may require the discovery of new physics, meaning that whatever the true value is, it is not one that is known currently. The problem then becomes one of deciding how to implement Douven's model in the case that the true value of τ is not known.

Following the work of Frey and Šešelja, I will address this challenge through the method of *robustness analysis*.[4] The authors argue that an ABM provides evidence for philosophical and historical hypotheses only if it is robust under two types of changes: changes in the parameters used and changes in the idealizing assumptions of the model. I address both types of changes in my ABM implementation below and offer a *rationale* for how these changes model the Hubble tension realistically. My implementation consists of an observation stage and a conference stage.

3.1 The Observation Stage

Table 6 in Appendix A lists the eight most recent measurements of H_0.[5][6][7][8][9][10][11] The goal of the observation stage is to model how each of the collaborations arrived at its respective value of H_0. This is implemented by running simulations of Douven's ABM under the choice of parameters uniquely characterizing each collaboration. Doing so requires changing the values of n, τ, and ζ from those listed in Table 1. The value of n is set to the size of each collaboration as inferred from the number of authors in each collaboration. The values of τ and ζ have been respectively inferred from the published values of H_0 and their error bars given in Table 6. The rationale behind this choice is that each value of τ should not represent the true value of H_0, which is unknown, but rather the value determined given each collaboration's *methodology*. Conveniently, this offers a straightforward interpretation of the meaning of ζ: the noise encountered by agents is none other than the sources of error of a collaboration's methodology. Thus, each collaboration's τ value corresponds to their published H_0 value in the literature.

Finally, Douven's model assigns his agents an initial belief drawn randomly from the uniform distribution $[0, 1]$. Instead, I've assigned my agents a value drawn from the $[64, 80]$ distribution in agreement with the first precise Hubble constant value of 72 ± 8 km s^{-1} Mpc^{-1} measured by Freedman et al.[12] The idea behind this is to model scientists who have formed collaborations with the intent of improving upon an initial measurement. I will run 100 simulations of 50 updates for each collaboration using the parameters of Table 2 to compare results with and without difference-splitting.

[4] Frey and Šešelja 2018.
[5] Planck Collaboration et al. 2018.
[6] Abbott et al. 2018.
[7] Riess et al. 2019.
[8] Freedman, Madore, Hatt, et al. 2019.
[9] Huang et al. 2018.
[10] Wong et al. 2019.
[11] The Megamaser Cosmology Project (MCP) and Surface Brightness Fluctuations (SBF) values were presented in a conference, see (Verde, Treu, and Riess 2019)
[12] Freedman, Madore, Gibson, et al. 2001.

3.1.1 Expectations

My first expectation is that the simulations here won't differ significantly from Douven's results. The reason for this is that assigning a value to τ and ζ outside the unit interval does not have a direct effect on the update rule given in (3). This means that the agents should be able to converge to their respective τ value within a finite number of updates. Therefore, I expect that Douven's main findings will be replicated: no difference-splitting simulations will rapidly get close to τ without further improvement whereas difference-splitting simulations will take longer to get close to the correct value of τ but, once they do, they will get much closer.

My second expectation is that both difference-splitting and no difference-splitting simulations will have smaller average least-squares penalties compared to Douven's results. I base this expectation on the fact that in Douven's simulations the ratio of ζ to τ is $0.2/0.75 \approx 27\%$, which is substantially larger than the same ratio computed for each collaboration. Because the same calculation ranges from 0.7% for the Planck collaboration to 5% for the SBF collaboration, I expect all least-squares penalties to be smaller.

Now, the biggest difference among the collaborations outside of their respective τ and ζ values is in the number of agents. SBF, the smallest collaboration, has 8 authors while Planck, the largest collaboration, has 178 authors. One might be tempted to think that such a wide distribution in the number of agents might have an effect on the simulations. However, condition (2) pools the opinions of *all* agents falling within the tolerance interval irrespective of size. Therefore, my third expectation is that because size effects have no effect on the update rule. agents will converge towards a consensus at the same number of updates for each collaboration.

3.2 The Conference Stage

While the previous modification of Douven's ABM aimed at presenting a robustness analysis under parameter changes, this modification of the ABM changes the idealizing assumptions of the model. Scientists in the conference stage are no longer taking measurements, but rather communicating their results to each other. Thus, there is no τ value in this stage of the model and agents will change their beliefs only through opinion pooling. The conference simulations will be initialized with 100 agents selected randomly from all collaborations with their initial value set to their last believed value of H_0 at the end of the observation stage. The rationale for this modification is that, rather than arriving at the conference with random beliefs, agents have a well-defined belief of H_0 based on their individual measurements leading to the publication of their collaboration's results.

A further modification is that agents no longer have the same size of tolerance interval ε. Instead, condition (2) is replaced by

$$Z_i(u) := \{j : |x_i(u) - x_j(u)| \leq \zeta_i\} \tag{4}$$

to determine which of their peers agents are willing to listen to at the conference. Here, $Z_i(u)$ differs from $X_i(u)$ in the sense that $\varepsilon = \zeta_i$ where the index in ζ_i corresponds to the uncertainty in the result published by agent i's collaboration. Thus, agents will be open to changing their minds only insofar as they are willing to listen to agents whose reported values fall within the noise level they encountered during the observation stage.

As discussed in Appendix A, measurements of the Hubble constant can be classified in terms of whether they were obtained from observations of the early universe or from observations of the late universe. Agents may be implicitly biased towards agents from collaborations employing similar methodologies and to account for this, two types of opinion pooling will be considered: one in which agents listen to all other agents fairly and another in which agents are biased against agents who took measurements of an observation epoch different to their own.

The update rule for the unbiased case is given by

$$x_i(u+1) = \frac{1}{|Z_i(u)|} \sum_{j \in Z_i(u)} x_j(u) \tag{5}$$

On the other hand, the update rule for the biased case is given by:

$$x_i(u+1) = \frac{\beta}{|M_i(u)|} \sum_{j \in M_i(u)} x_j(u) + \frac{(1-\beta)}{|\neg M_i(u)|} \sum_{j \in \neg M_i(u)} x_j(u) \tag{6}$$

Here β is a real number between 0 and 1 playing the role of weighing opinions similar to α, but in this case, the weighing is done in terms of what observation epoch other agents fall into. In this paper, $\beta = 0.9$ will be held fixed to specify strong bias. Defining the characteristic function

$$m(x_i) = \begin{cases} 1 & \text{if } x_i \text{ took early universe measurements} \\ 0 & \text{if } x_i \text{ took late universe measurements} \end{cases}, \tag{7}$$

the set

$$M_i(u) := \{j \in Z_i(u) : m(x_j) = m(x_i)\} \tag{8}$$

consists of agents whose reported values fall within the i-th agent's tolerance interval and took measurements from the same observation epoch, whereas the set

$$\neg M_i(u) := \{j \in Z_i(u) : m(x_j) \neq m(x_i)\} \tag{9}$$

consists of agents who, although reporting values within agent i's tolerance interval, took measurements from a different observation epoch. These are the agents that will be penalized by the update rule.

3.2.1 Expectations

Unlike the changes in the parameter space implemented in the observation stage, the conference stage involves significant changes in the idealizing assumptions of Douven's model. My first expectation is that because there is no definite τ value in the update rule, the agents will fail to converge around a unique value. To qualify this expectation, let's take a look at the initial conditions. Out of 362 total agents, 178 correspond to the Planck collaboration and 96 correspond to the DES collaboration, meaning that 75.7% of the sample space consists of early universe agents. Because the Planck collaboration has the smallest tolerance interval, its agents will be too stubborn to listen to the early universe agents, thereby preventing the conference stage from converging around a single τ value.

My second expectation is that in the unbiased conference, the late universe agents will be substantially attracted to the beliefs of the early universe agents. The reasoning behind this is that the late universe agents have larger tolerance intervals and a much larger range in their initial H_0 values. Moreover, they correspond to only 24.3% of the sample space, so it seems unlikely that they would attract all of the early universe agents to their beliefs. Instead, any late universe agents interacting with early universe agents will be drawn to their beliefs, while the early universe agents will persist to ignore their input.

My third expectation pertains to how the results for the biased conference will differ from those of the unbiased conference. Instead of taking a fair average of all agents falling within their tolerance interval, agents will now strongly favor agents in the opinion pool who took measurements of the same epoch ($\beta = 0.9$) and disfavor agents in the opinion pool who took measurements of a different epoch ($\beta = 0.1$). Thus, I expect the segregation of early universe agents to intensify to the point that they will listen almost exclusively to each other. At the same time, late universe collaborations will give less weight to the opinion of the early universe collaborations, allowing them to arrive at their own consensus value independently from that of the early universe collaborations.

4 Discussion

The plots for all simulation results may be found in Appendix B. Using these, we can now infer how well my expectations were met.

4.1 Observation Stage

My first expectation for the observation stage was that the main results of Douven's simulations were going to repeat themselves: no difference-splitting communities would approach the true value of τ really fast without further improvement while difference-splitting communities would take longer to get close to τ but, upon doing so, they would get much closer to the correct value. I found a really good agreement with this prediction for all eight collaborations. The agreement is readily discernible in Figures 1 through 16 and is also reproduced in Table 3 where the least-squares penalties at the fifth, tenth, twenty-fifth, and last update for each collaboration, averaged over 100 simulations, are displayed.

My second expectation for the observation stage was that, because the ratios of the collaborations' errors to their τ value were much smaller than in Douven's simulations, the size of the penalties would be smaller accordingly. This time my expectation was not met. Comparison to Douven's Table 2 shows that none of the least-squares penalties displayed in my Table 3 are smaller than Douven's values. Puzzled by this, I ran simulations using Douven's parameters and was able to reproduce his Table 2 results exactly.[13] I've since concluded that my simulations have larger least-squares penalties due to the larger spread in the agents' initial beliefs. Note, for example, the very high penalties at the fifth update in the difference-splitting column of Table 3, which can be explained

[13]Douven 2010, p.152.

from the fact that all agents were randomly initiated by drawing from the $[64, 80]$ interval.

The third and last expectation was that collaboration size should have no effect on how many updates it takes communities to converge around their correct τ value. For the difference-splitting case, all collaborations improved steadily irrespective of their size. However, this expectation was not met by the Miras, MCP, and SBF collaborations in the no difference-splitting case. As seen in the first column of Table 3, none of these collaborations saw substantial improvement in their least-squares penalties. This poor performance should not be attributed to the small collaboration size, but rather to their having encountered the largest noise levels. For all other collaborations, the no difference-splitting results of Douven's Table 2 are replicated: the least-squares penalties are most significantly reduced at the 10-th update without further improvement.

Now, Frey and Šešelja write that we can speak of the robustness of results under parameter changes by specifying the scope of parameters within which results remain stable enough to display the phenomenon under consideration.[14] As the above discussion shows, the main difference between difference-splitting and no difference-splitting communities are preserved under parameter changes outside the unit interval, but the value and time to convergence of the agents' least-squares penalties are found to be highly sensitive to the parameter choices used to model a given scientific collaboration. The fact that the simulations were nonetheless able to converge within a very small range from the correct τ value—a range smaller than the error bar of each collaboration—attests to the robustness of the model under parameter changes drawn outside the unit interval.

4.2 Conference Stage

We now turn to a discussion of the results for the conference stage, presented in Figures 17 through 23. Table 4 summarizes the results of 100 simulations for the unbiased conference and the biased conference. The first column describes how many distinct end-states remained after the last update while the second and third column describe the relative frequency of simulations ending with that many states. Already we can see that my first expectation was not met: at least one simulation terminated with all agents converging to a single unique value of H_0, shown in Figure 17. However, this is very unusual, as only 1 out of 100 simulations of the unbiased conference produced this result.[15]

My second expectation was that the late universe agents in the unbiased conference would be substantially attracted to the beliefs of the early universe agents. This expectation was also not met. Figure 18 shows a case in which the late universe reach a consensus among themselves with no further interactions among the agents. As suggested by Table 4, such two end-state scenarios occur about 51% of the time. Similarly, Figure 19 shows a scenario in which the CCHP agents, those initialized with a τ value around 69.8, do not interact with anyone outside their collaboration. Although this scenario was not unique among the three end-state cases (as the examination of other plots not here

[14] Frey and Šešelja 2018, p.415.

[15] Moreover, other 100 simulation runs resulted in cases where none of the agents converged to a unique end-state.

depicted revealed), it still lends support to the observation that late universe agents were not ubiquitously attracted to the early universe agents.

The third expectation was that the segregation effect between early and late universe agents would intensify in the biased conference as compared to the unbiased conference. The main qualitative difference between the unbiased conference figures and the biased conference figures is that the early universe agents in the latter never deviated from their consensus. Table 4 shows that 98% of the biased conference simulations ended with two or three distinct end-states. Thus, the scenario of Figure 23 in which the CCHP agents do not interact with anyone else seems to be significantly rarer than in the unbiased conference stage. This being the case, the results here presented suggest that my third expectation was met and segregation between early universe agents and late universe agents intensified.

Overall, the main result to be ascertained from the conference stage is that early universe agents are very stubborn. This effect is best represented in Table 5, where I took the mode of each simulation and the percentage of agents having that value before averaging across all simulations. The results for the unbiased and biased conference have a near negligible deviation from the initial distribution of 75.7% of the agents believing in a H_0 value of 67.4 km s^{-1} Mpc^{-1}. This means that notwithstanding the great variability in the opinion dynamics witnessed in Figures 17 through 23, most of the early universe agents in fact never interacted with agents outside of their collaboration.

Collaboration	No difference-splitting	Difference-splitting
Planck	$\langle 0.3892, 0.0556, 0.0561, 0.0557\rangle$	$\langle 36.3486, 12.6704, 0.5316, 0.0028\rangle$
DES	$\langle 0.6377, 0.2970, 0.2909, 0.2902\rangle$	$\langle 36.5649, 12.7654, 0.5615, 0.0165\rangle$
SH0ES	$\langle 0.6574, 0.4342, 0.4401, 0.4431\rangle$	$\langle 22.4840, 7.9394, 0.3835, 0.0459\rangle$
CCHP	$\langle 1.0570, 0.8043, 0.8012, 0.8406\rangle$	$\langle 22.5798, 7.9083, 0.4525, 0.1041\rangle$
Miras	$\langle 3.6468, 3.3744, 3.5019, 3.5316\rangle$	$\langle 21.2561, 7.9157, 0.8702, 0.5202\rangle$
H0LiCOW	$\langle 0.8578, 0.6714, 0.6748, 0.6735\rangle$	$\langle 18.5736, 6.4954, 0.3637, 0.0765\rangle$
MCP	$\langle 2.3113, 2.1214, 2.1687, 2.1049\rangle$	$\langle 24.7712, 8.9471, 0.6811, 0.3181\rangle$
SBF	$\langle 3.7296, 3.2846, 3.6359, 3.5348\rangle$	$\langle 36.8486, 13.3117, 1.0767, 0.5249\rangle$

Table 3: Average least-squares penalties for the observation stage after the fifth, tenth, twenty-fifth and last update (100 simulations, rounded to four decimal places)

Distinct end-states	Unbiased	Biased
1	1	0
2	51	49
3	35	49
4	13	2

Table 4: Relative frequencies of distinct end-states after 100 simulations of the conference stage.

	Unbiased	Biased
Average mode	67.5 km s^{-1} Mpc^{-1}	67.4 km s^{-1} Mpc^{-1}
Average percentage	74.3%	74.8%

Table 5: Average mode of the Hubble constant with relative frequency for after simulations of the conference stage.

Conclusion

My stated aim for this paper was to determine whether Douven's model of peer disagreement could be used to model the Hubble tension in cosmology. Insofar as the robustness analysis of the model is concerned, the results of my simulation's observation stage ensure that the model passed with flying colors. The simulations not only replicated the main findings for difference-splitting and no difference-splitting communities, but they also captured the salient features characterizing each collaboration. The implication of these findings for the history and philosophy of science is that it pays to choose model parameters from historically available data. As the science of physical cosmology develops in the years to come, the effects of collaboration size and methodological uncertainty can help both historians and philosophers to draw inferences on how research programs in astrophysics differ by observation epoch. Insofar as the results of the conference stage are concerned, I must reinstate how impressed I am by the richness of the simulations: I was here only able to discuss to what extent they met my initial expectations, but the great variability in the opinion dynamics certainly merits further discussion. In particular, it might be a good idea to repeat the biased and unbiased conference scenarios using Douven's default value of $\varepsilon = 0.1$ to see whether it results in more pronounced qualitative differences in the resulting plots.

I'd like to close by revisiting the status of the *Irrationality Claim* discussed by Douven. The main objection identified at the end of the first section was that changing Douven's simulations may produce results that no longer make peer disagreement so dependent on contingent factors. Although changes were indeed observed after introducing my modifications—especially in the conference stage where a fixed τ value was no longer available—these changes certainly reinforce the context-sensitivity of peer disagreement situations. The sensibility and high variability of the simulations here discussed supports the notion that the validity of IC cannot be established a priori and that instances of peer disagreement should be evaluated on a case-by-case basis. It is worth noting that although my conference stage simulations demonstrate how very stubborn early universe agents were in taking into account the results reported by their late universe peers, they do not imply that cosmologists would be justified in outright dismissing the contributions of late universe observations. The fact that they are starting to recognize that resolution of the Hubble tension may require the discovery of new physics beyond the Standard Model of cosmology goes a long way to show that disagreement among scientists about physical measurements can be fully rational and be used to revise a community's epistemic goals. In this case, the truly irrational thing for cosmologists to believe would not be thinking that a resolution of the Hubble tension may require new physics, but rather thinking that such tension could be resolved *a priori* by blindly taking an average of all their measurements.

References

Abbott, T M C et al. (2018). "Dark Energy Survey Year 1 Results: A Precise H0 Estimate from DES Y1, BAO, and D/H Data". In: *Monthly Notices of the Royal Astronomical Society* 480.3, pp. 3879–3888.

Carroll, Bradley W and Dale A Ostlie (2007). *An introduction to modern astrophysics; 2nd ed.* San Francisco, CA: Addison-Wesley.

Douven, Igor (2010). "Simulating Peer Disagreements". In: *Studies in History and Philosophy of Science Part A* 41.2, pp. 148–157.

Freedman, Wendy L., Barry F. Madore, Brad K. Gibson, et al. (2001). "Final Results from the Hubble Space Telescope Key Project to Measure the Hubble Constant". In: *Astrophysics Journal* 553.1, pp. 47–72.

Freedman, Wendy L., Barry F. Madore, Dylan Hatt, et al. (2019). "The Carnegie-Chicago Hubble Program. VIII. An Independent Determination of the Hubble Constant Based on the Tip of the Red Giant Branch". In: *The Astrophysics Journal* 882.1, 34, p. 34.

Frey, Daniel and Dunja Šešelja (2018). "What Is the Epistemic Function of Highly Idealized Agent-Based Models of Scientific Inquiry?" In: *Philosophy of the Social Sciences* 48.4, pp. 407–433.

Huang, Caroline D. et al. (2018). "A Near-infrared Period-Luminosity Relation for Miras in NGC 4258, an Anchor for a New Distance Ladder". In: *The Astrophysics Journal* 857.1, 67, p. 67.

Planck Collaboration et al. (2018). "Planck 2018 results. VI. Cosmological parameters". In: *arXiv e-prints*.

Riess, Adam G. et al. (2019). "Large Magellanic Cloud Cepheid Standards Provide a 1% Foundation for the Determination of the Hubble Constant and Stronger Evidence for Physics beyond ΛCDM". In: *The Astrophysics Journal* 876.1, 85, p. 85.

Smeenk, Christopher and George Ellis (2017). "Philosophy of Cosmology". In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Winter 2017. Metaphysics Research Lab, Stanford University.

Verde, Licia, Tommaso Treu, and Adam G. Riess (2019). "Tensions between the early and late Universe". In: *Nature Astronomy* 3, pp. 891–895.

Wong, Kenneth C. et al. (2019). "H0LiCOW XIII. A 2.4% measurement of H_0 from lensed quasars: 5.3σ tension between early and late-Universe probes". In: *arXiv e-prints*.

Appendix A: The Hubble Constant

This appendix introduces the relevant concepts and history necessary to understand the Hubble constant and appreciate its significant to physical cosmology. In particular, we will distinguish between values inferred from observations of the late universe from those inferred from observations of the early universe which, while nowhere near comprehensive enough to permit a rigorous analysis of the methodologies involved, should be enough to stress that the Hubble tension emerges as the discrepancy between measurements inferred from two distinct observation epochs.

The theoretical basis for physical cosmology consists of solving the Einstein Field Equations to obtain a metric tensor describing the space-time geometry of the whole universe.[16] These equations are not solved directly, but rather greatly simplified from their original form by assuming that the universe is homogenous and isotropic. This assumption reduces the Einstein Field Equations to three coupled differential equations:

The Friedmann Equation

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G\rho}{3} - \frac{k}{a^2} + \frac{\Lambda}{3}, \tag{10}$$

The isotropic Raychaudhuri Equation

$$3\frac{\ddot{a}}{a} = -4\pi G(\rho + 3p) + \Lambda, \tag{11}$$

and the conservation of energy for a perfect fluid.

$$\dot{\rho} + (\rho + p)3\frac{\dot{a}}{a} = 0 \tag{12}$$

The solution of these equations results in the Friedmann–Lemaître–Robertson–Walker metric describing the space-time geometry of the universe:

$$ds^2 = a(t)^2 ds_3^2 - c^2 dt^2 \tag{13}$$

Crucially, the function $a(t)$ describes how every region of space evolves with respect to time. Known as the scale factor, its behavior is governed jointly by the three equations above. For the purposes of this paper, the most important theoretical detail to keep in mind is that the Hubble constant is defined as the ratio of the time derivative of the scale factor to itself:

$$H_0 \equiv \frac{\dot{a}(t)}{a(t)} \tag{14}$$

Finding the value of H_0, therefore, corresponds to finding a parameter that describes the entire expansionary history of the universe.

The first observational evidence for the expansion of the universe was concurrent with the discovery of the existence of galaxies beyond the Milky Way. Edwin Hubble's 1925 discovery of Cepheid variable stars in M31 settled the Great Debate concerning whether the "spiral nebulae" observed by astronomers

[16] Smeenk and Ellis 2017.

were smaller structures at the edge of the Milky Way or their very own self-contained "island universes."[17] Cepheid variables are periodic pulsating stars. By measuring the change in brightness of the star's surface as it expands and contracts, astronomers are able to infer a period-luminosity relation from which they can calculate the distance to the star. Hubble's measurements of Cepheids in $M31$ verified that they were indeed too far to be part of the Milky Way, thereby confirming the existence of galaxies external to our own. He then supplemented these measurements with redshift measurements of the stars' host galaxies to infer a relation between the galaxies' recessional velocity and their distance from Earth. By 1929 Hubble had measurements of 18 galaxies, which led him to the startling conclusion that a galaxy's recessional velocity v is proportional to its distance d. This relation,

$$v = H_0 d, \tag{15}$$

called Hubble's Law, describes a linear relation between the recessional velocity of the galaxy in km s^{-1} and its distance in megaparsecs. The units of the Hubble constant are, therefore, given in dimensions of km s^{-1} Mpc^{-1}. We thus see that the Hubble constant is not only highly theoretically motivated, it also satisfies an empirical law at the heart of modern astronomy.

Hubble estimated the original value of H_0 at around 500, but by the end of the 20th Century, astronomers had realized that it had to be much smaller, placing the value at somewhere in the $[50, 90]$ interval. The first precise measurement of the value was obtained by Freedman et al. in 2001.[18] Using Hubble's namesake telescope, the collaboration measured a Hubble constant value of 72 ± 8 km s^{-1} Mpc^{-1}. Since then, measurements have become increasingly more precise and have begun to rely on techniques other than distance-ladder measurements of Cepheid variables.

Key among these is the value obtained by the Planck collaboration from measurements of the cosmic microwave background (CMB). The CMB consists of left-over radiation from the end of the radiation-dominated era of expansionary history. Careful measurements of its spectrum allow cosmologists to solve the right-hand side of the Friedmann equation, resulting in a value for the Hubble constant that can then be extrapolated to infer the present rate of expansion. Because measurements of the CMB do not rely on direct measurements of nearby astrophysical sources, this determination of H_0 defines an early universe measurement statistically independent from late universe values inferred from distance ladder measurements. The problem today consists in understanding why both types of values are different.

On July 2019 approximately 100 cosmologists met at a conference held by the Kavli Institute for Theoretical Physics. The objective of the conference was to discuss the possible sources for the discrepancy between early and late universe measurements and to brainstorm a resolution. In addition, the most precise measurements of the Hubble constant to date were presented and assessed in terms of their relative strengths and weaknesses. Table 6 summarizes eight independent Hubble constant measurements discussed at the Conference.[19]

[17] Carroll and Ostlie 2007, p.1052.
[18] Freedman, Madore, Gibson, et al. 2001.
[19] Verde, Treu, and Riess 2019.

The main point to take away from this appendix is that, whatever the true value of the Hubble constant is, it is unlikely that it is one of the values published. After all, if this were the case and say the CCHP's value happens to be the correct one, it would mean that every other measurement was wrong. However, the scientists of each collaboration are very confident that they eliminated as much noise as possible from their methodology, which suggests that the resolution to the tension between their reported values may lie in new physics beyond the Standard Model of cosmology. If this is the case, then the correct value of the Hubble constant is most likely to be an intermediate value between those reported by the early universe collaborations and those reported by the late universe collaborations.

Collaboration	Number of authors	Observation Epoch	Methodology	Reported Value of H_0 (km s^{-1} Mpc^{-1})
Planck	178	Early	Early universe extrapolation from CMB spectrum	67.4 ± 0.5
DES	96	Early	Baryon acoustic oscillations and deuterium abundance	67.4 ± 1.15
SH0ES	15	Late	Distance-ladder calibration on Cepheid variables	74.0 ± 1.4
CCHP	13	Late	Distance-ladder calibration on tip of the red giant branch stars	69.8 ± 1.9
Miras	11	Late	Distance-ladder calibration on tip of AGB stars	73.6 ± 3.9
H0LiCOW	26	Late	Time-delay cosmography of gravitational lenses	73.3 ± 1.75
MCP	15	Late	Direct geometric distance measurements to circumnuclear H_2O megamasers	74.8 ± 3.1
SBF	8	Late	Surface brightness fluctuations of galaxies observed in the infra-red	76.5 ± 4.0

Table 6: Key parameters used in this paper's simulations

Appendix B: Simulation Plots

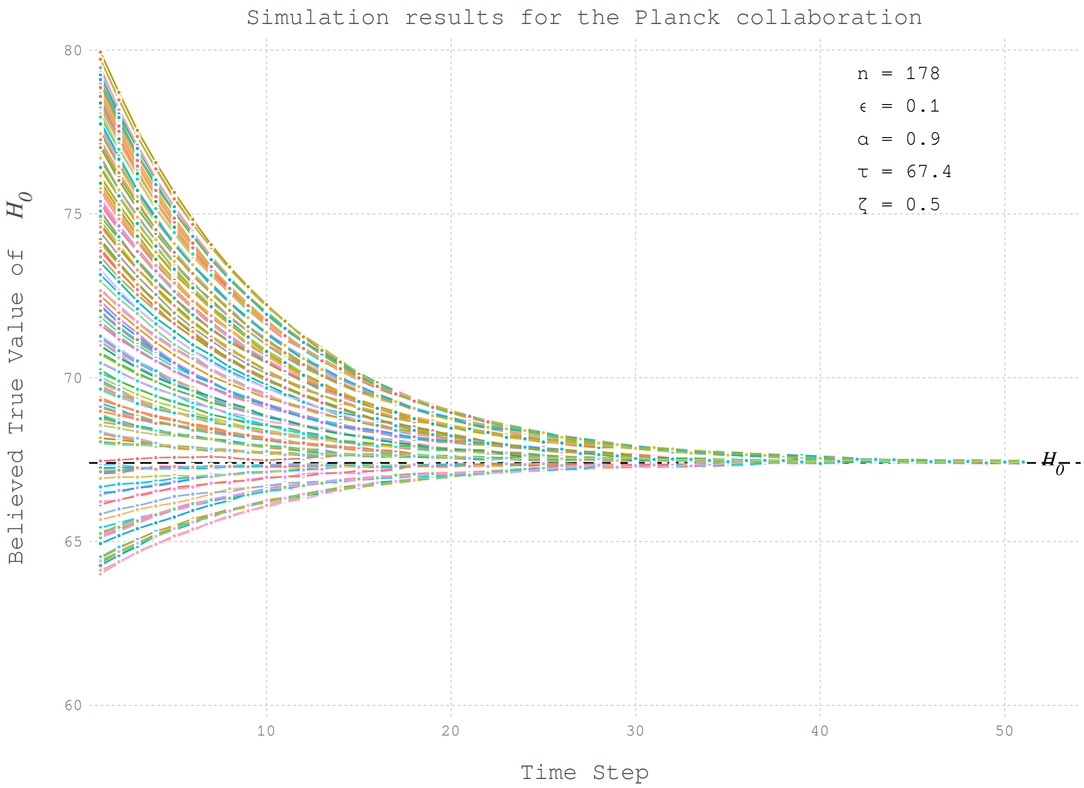

Figure 1: Planck collaboration with difference-splitting, showing the best convergence out of all collaborations.

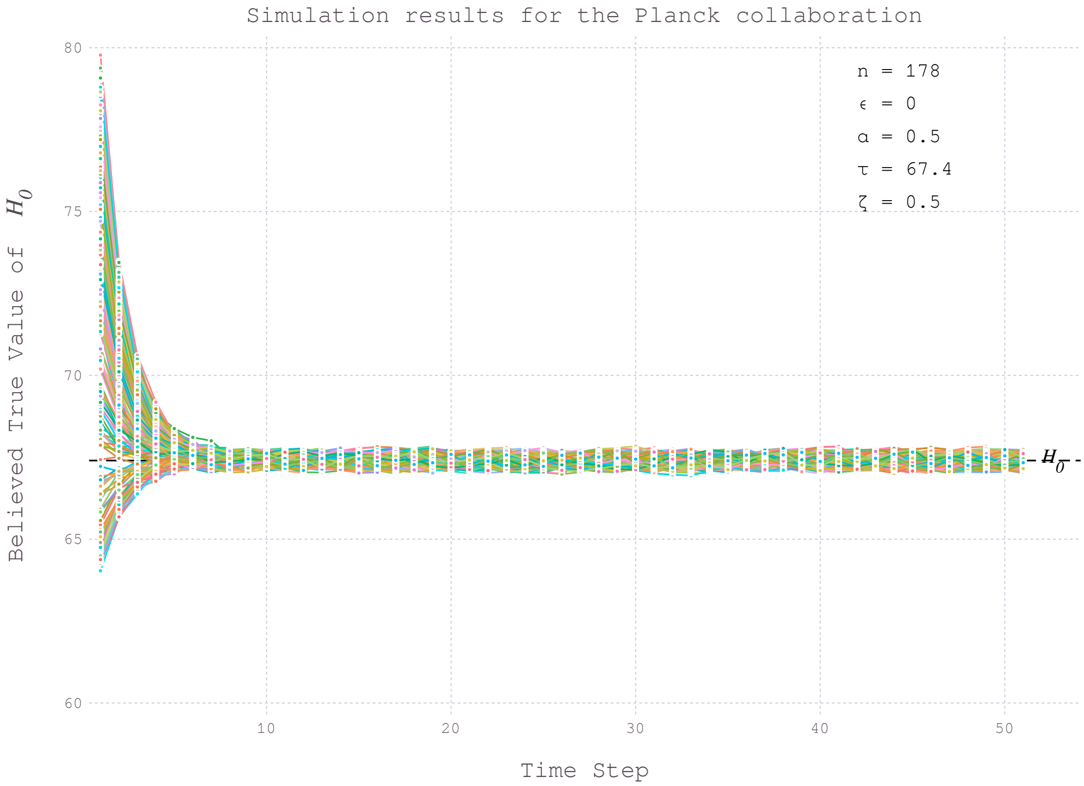

Figure 2: Planck collaboration with no difference-splitting.

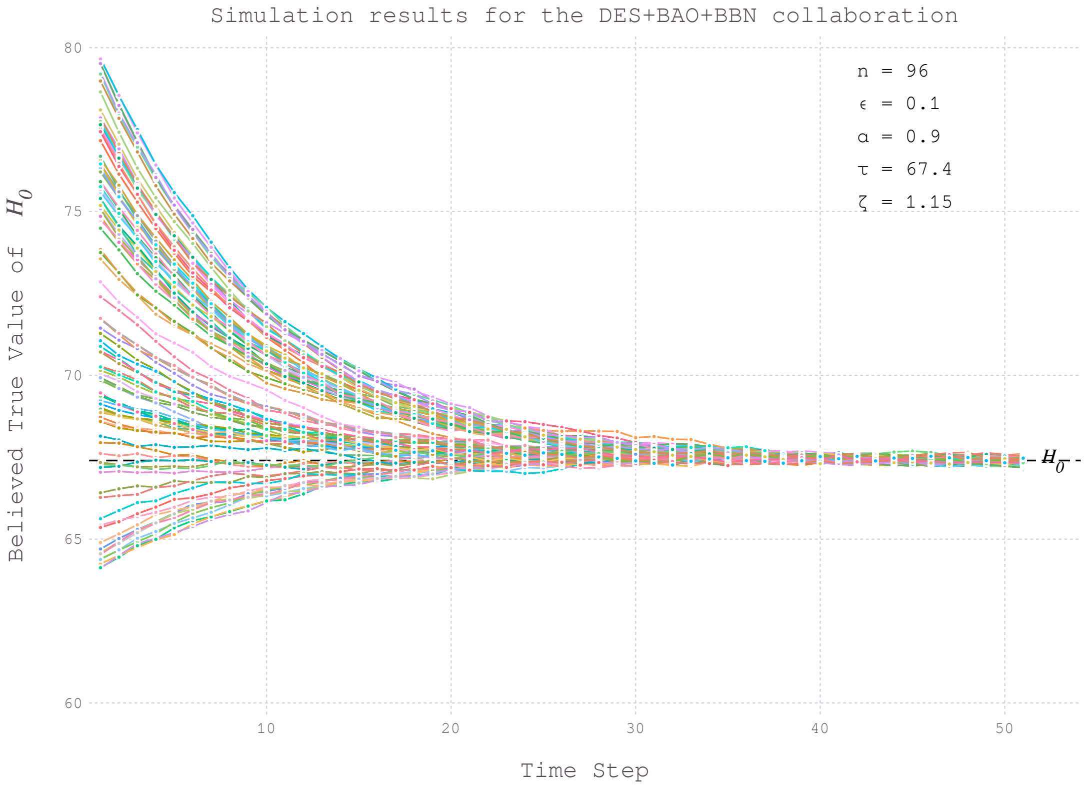

Figure 3: DES collaboration with difference-splitting, also showing very good convergence.

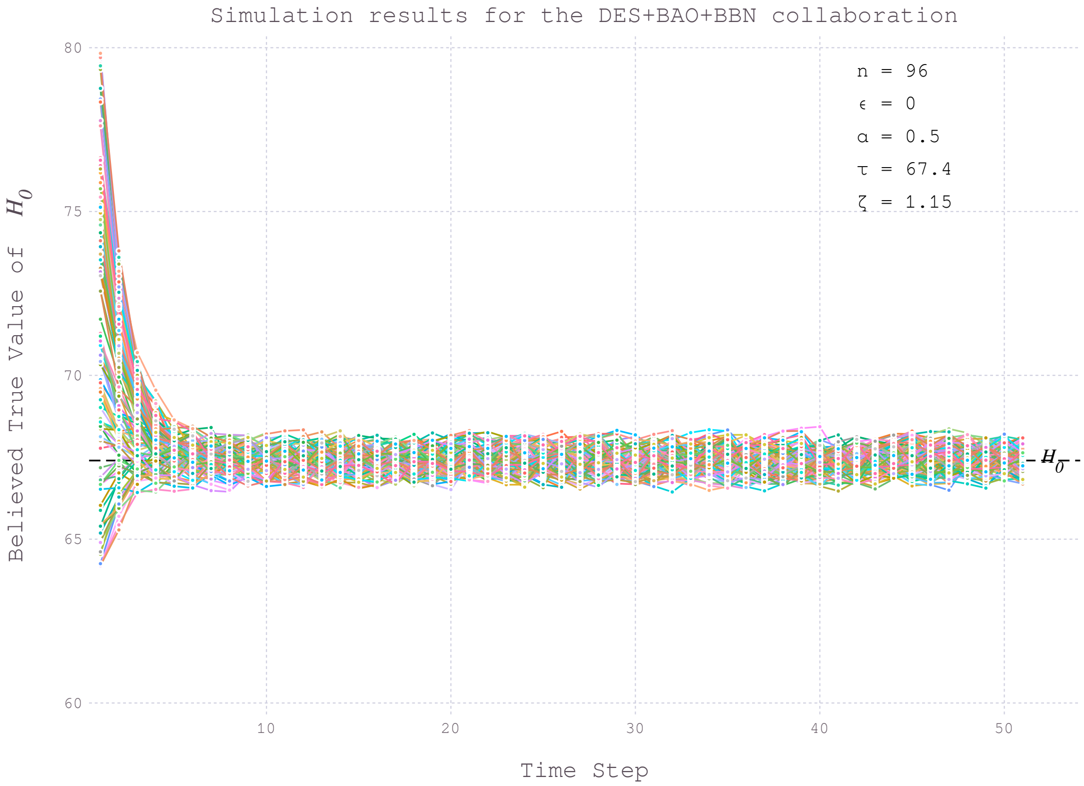

Figure 4: DES collaboration with no difference-splitting.

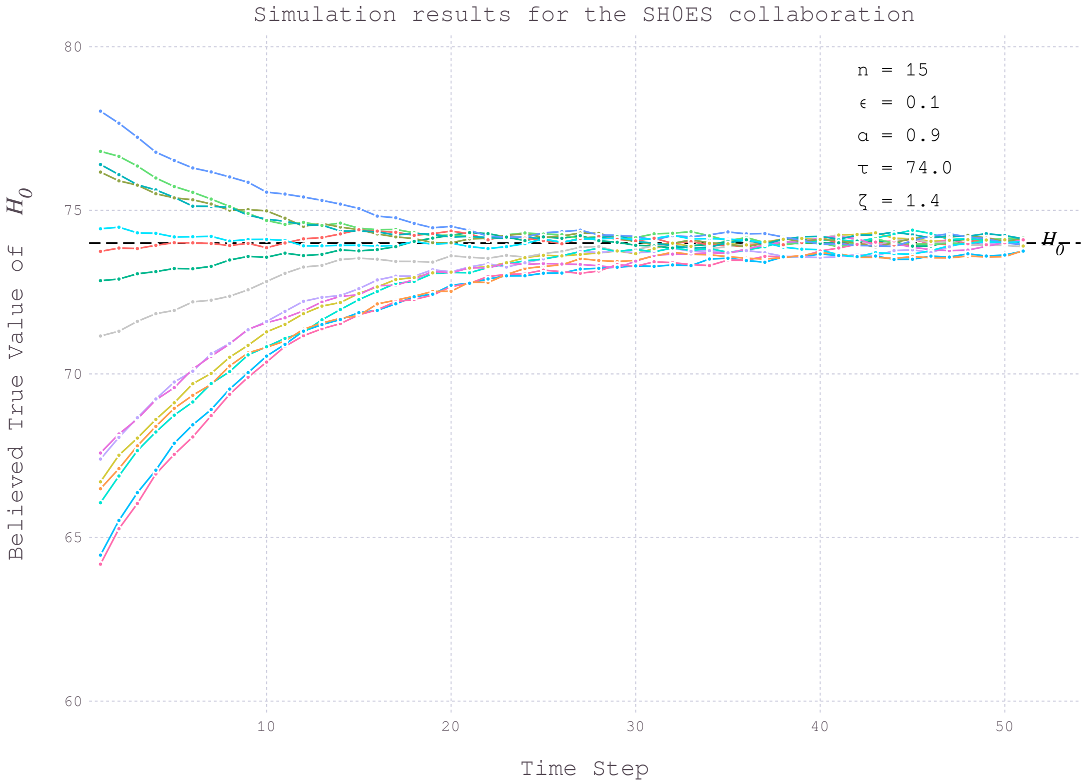

Figure 5: SH0ES collaboration with difference-splitting. Note how the convergence is not nearly as good as the early universe collaborations due to the larger noise levels.

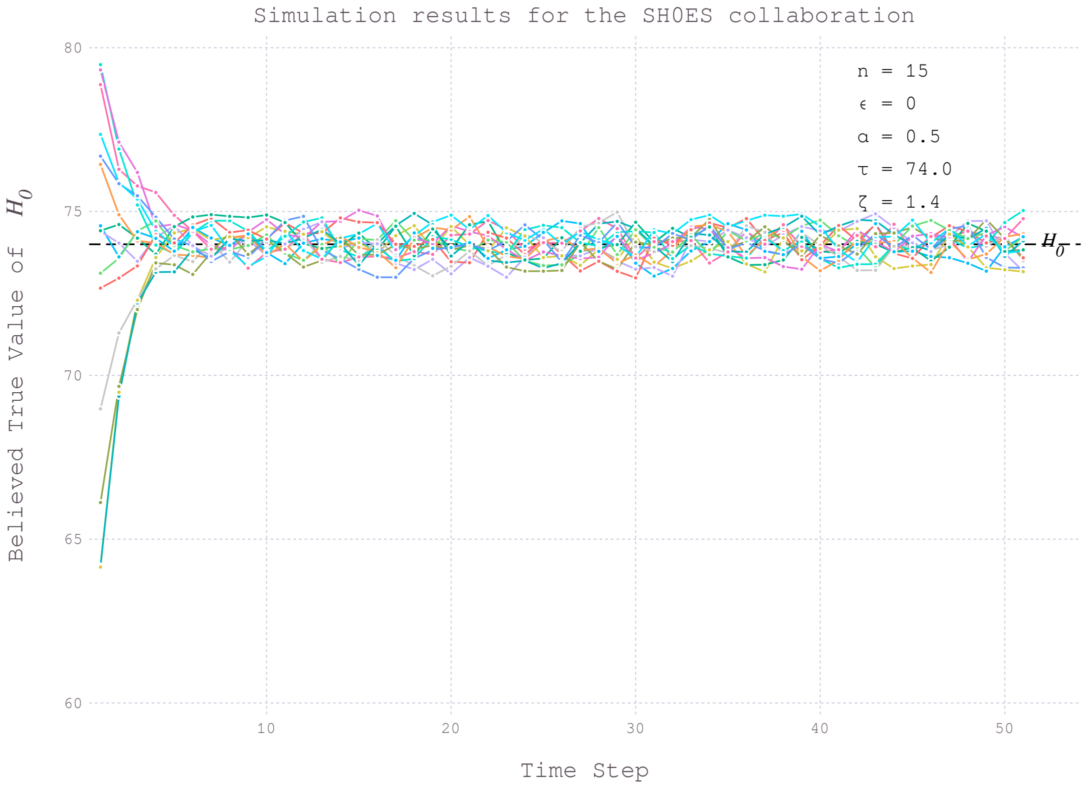

Figure 6: SH0ES collaboration with no difference-splitting.

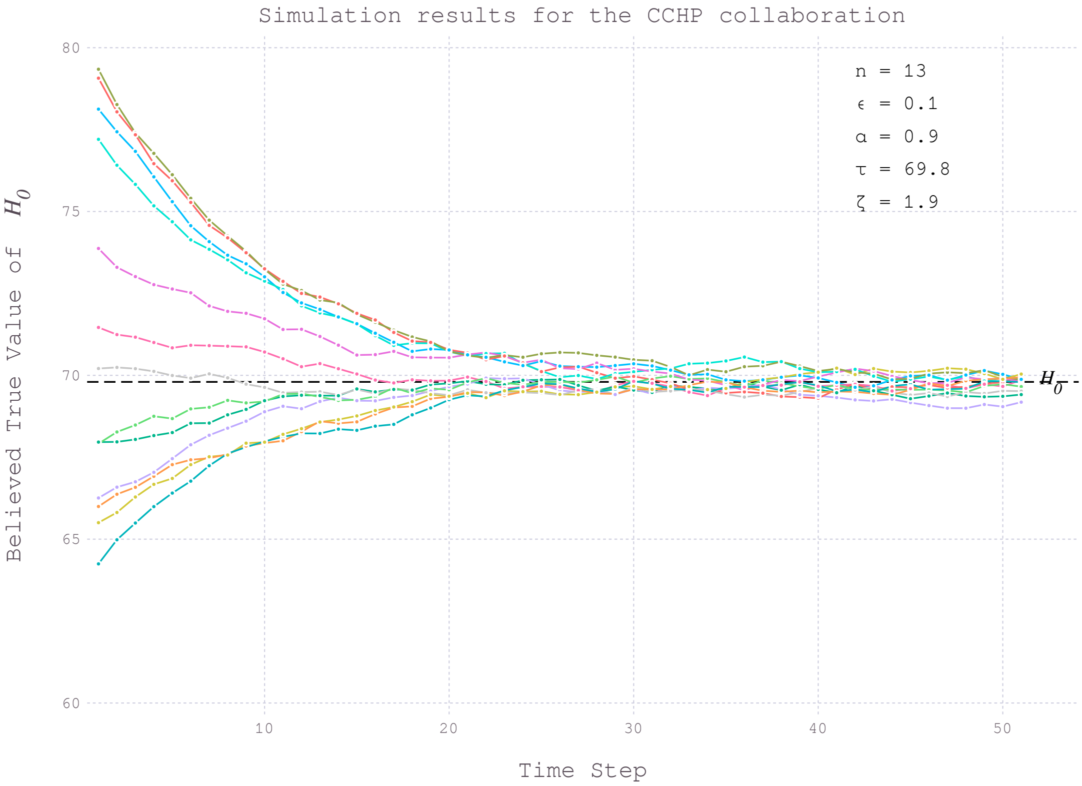

Figure 7: CCHP collaboration with difference-splitting.

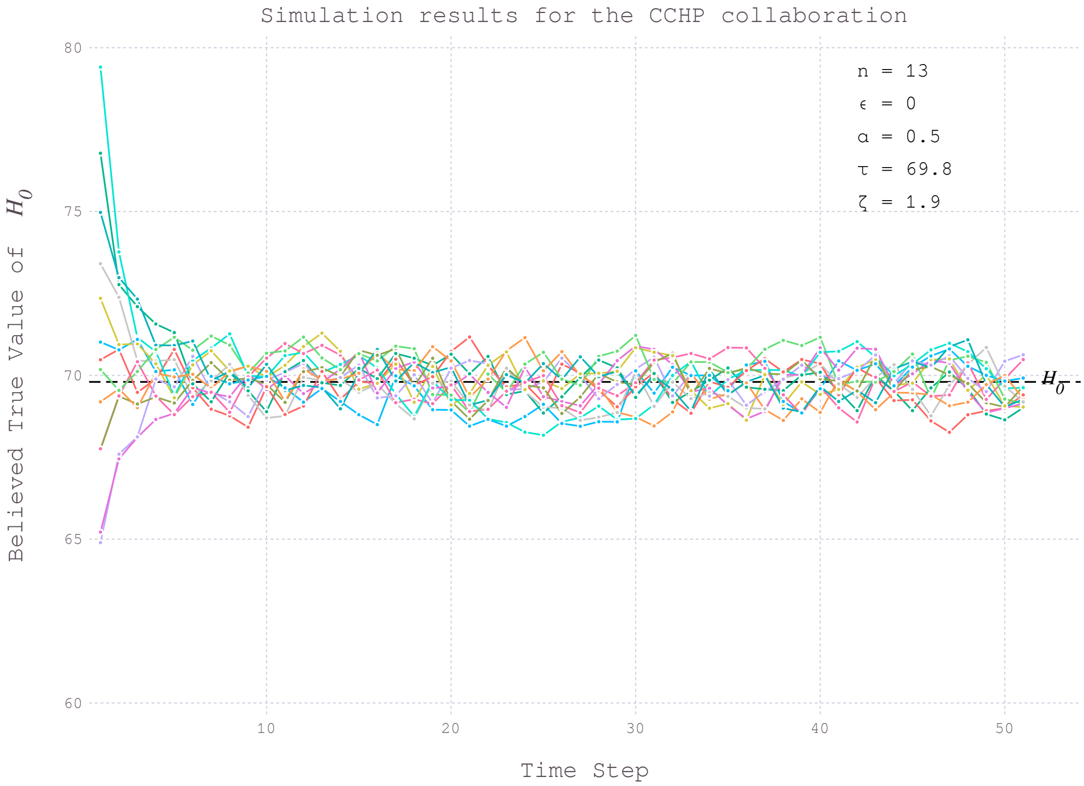

Figure 8: CCHP collaboration with no difference-splitting.

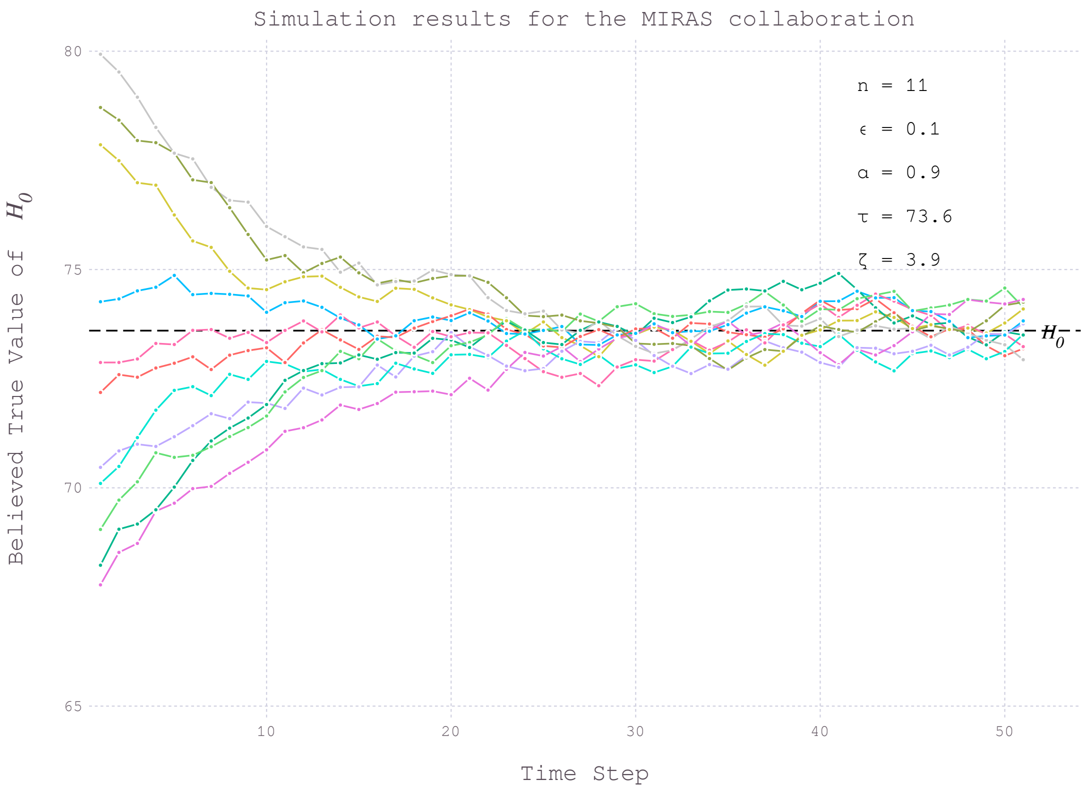

Figure 9: MIRAS collaboration with difference-splitting. Note the very poor convergence due to the great size of the noise level.

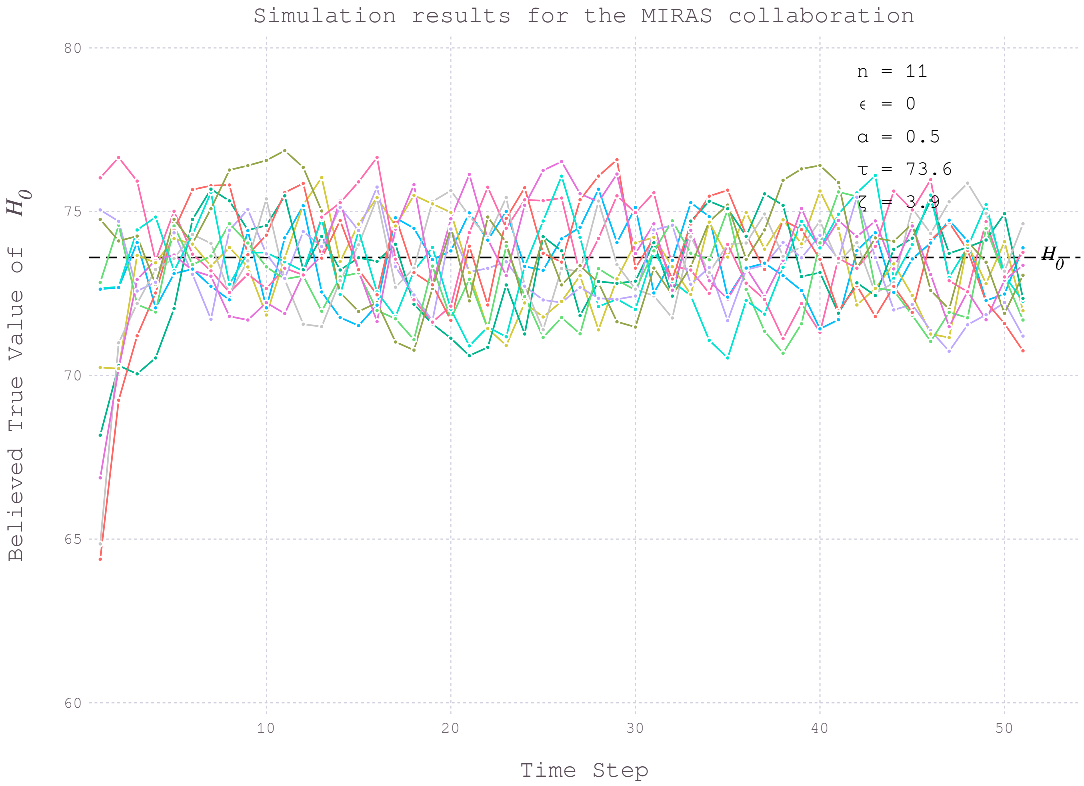

Figure 10: MIRAS collaboration with no difference-splitting, exhibiting even poorer convergence.

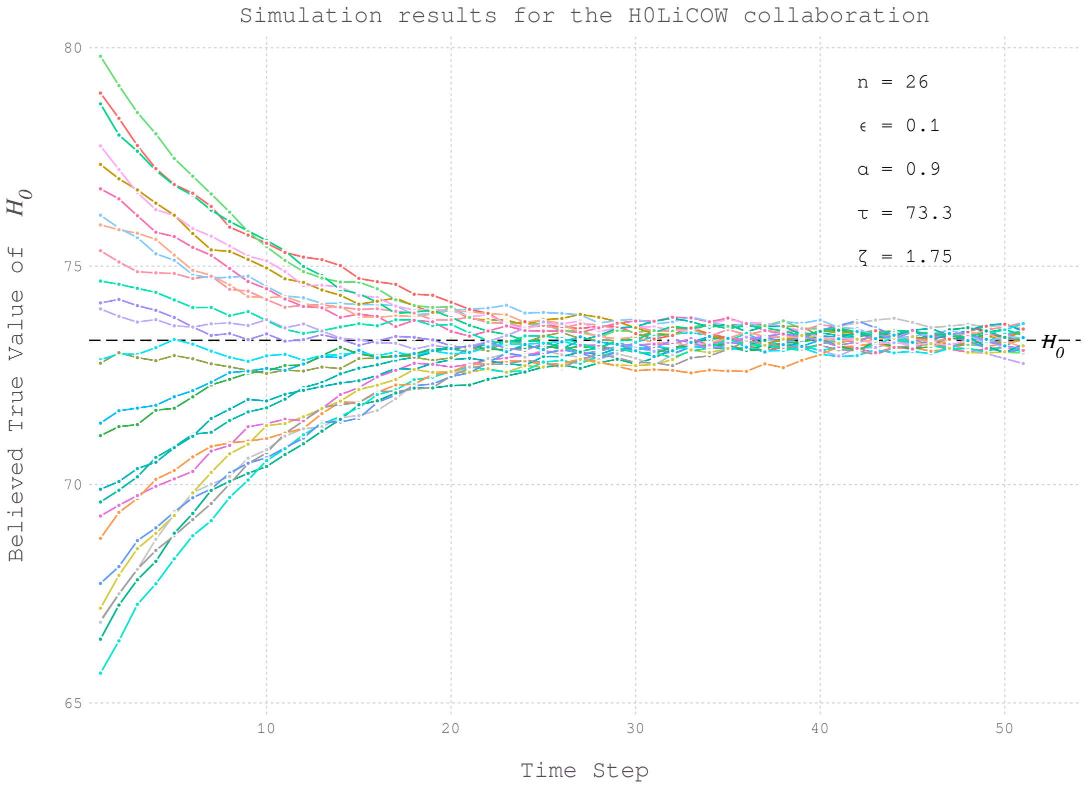

Figure 11: H0LiCOW collaboration with difference-splitting.

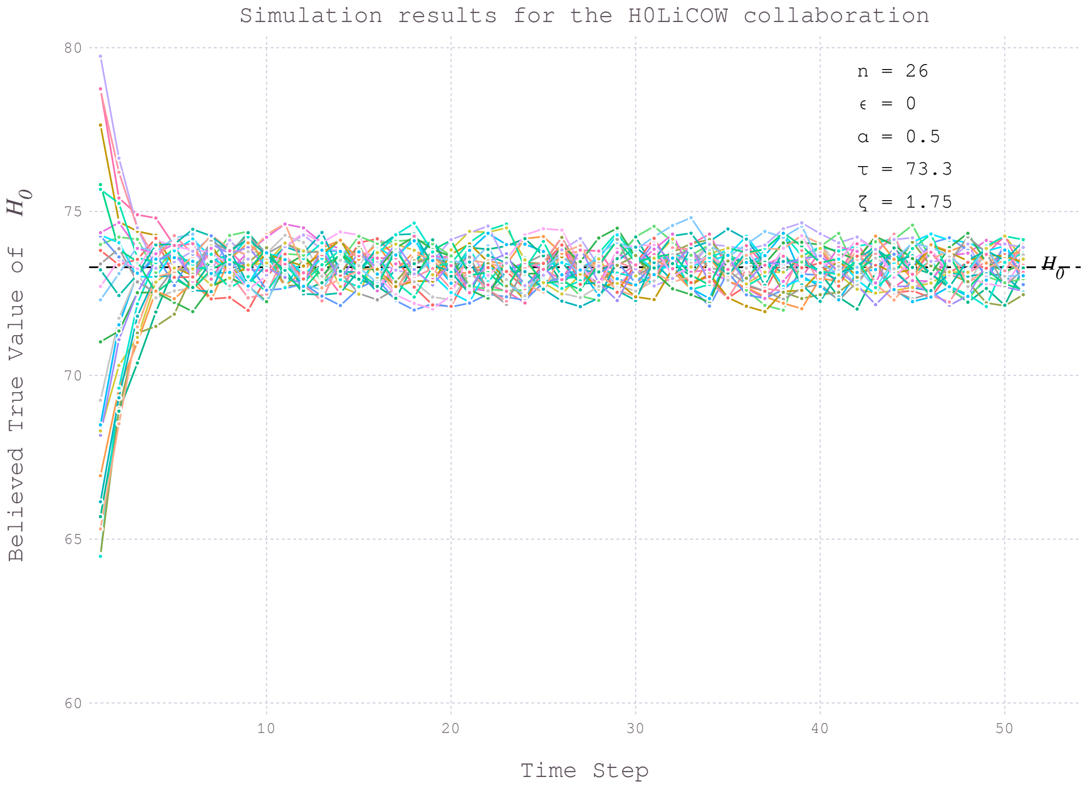

Figure 12: H0LiCOW collaboration with no difference-splitting.

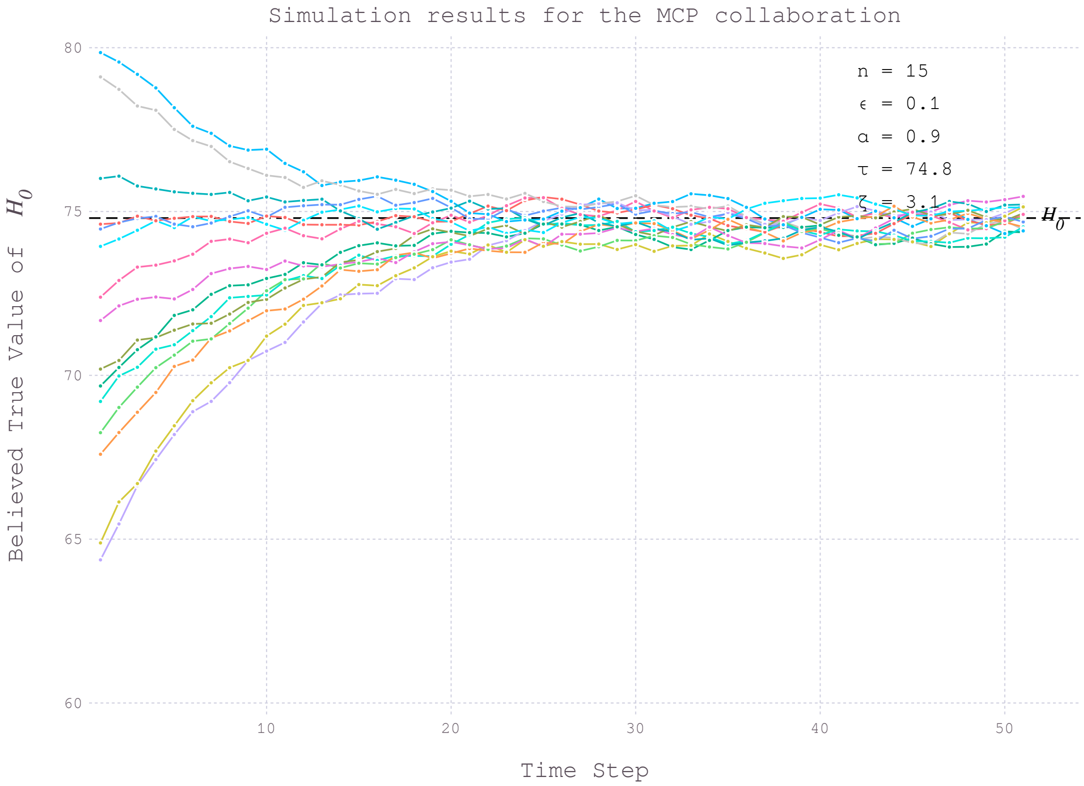

Figure 13: MCP collaboration with difference-splitting.

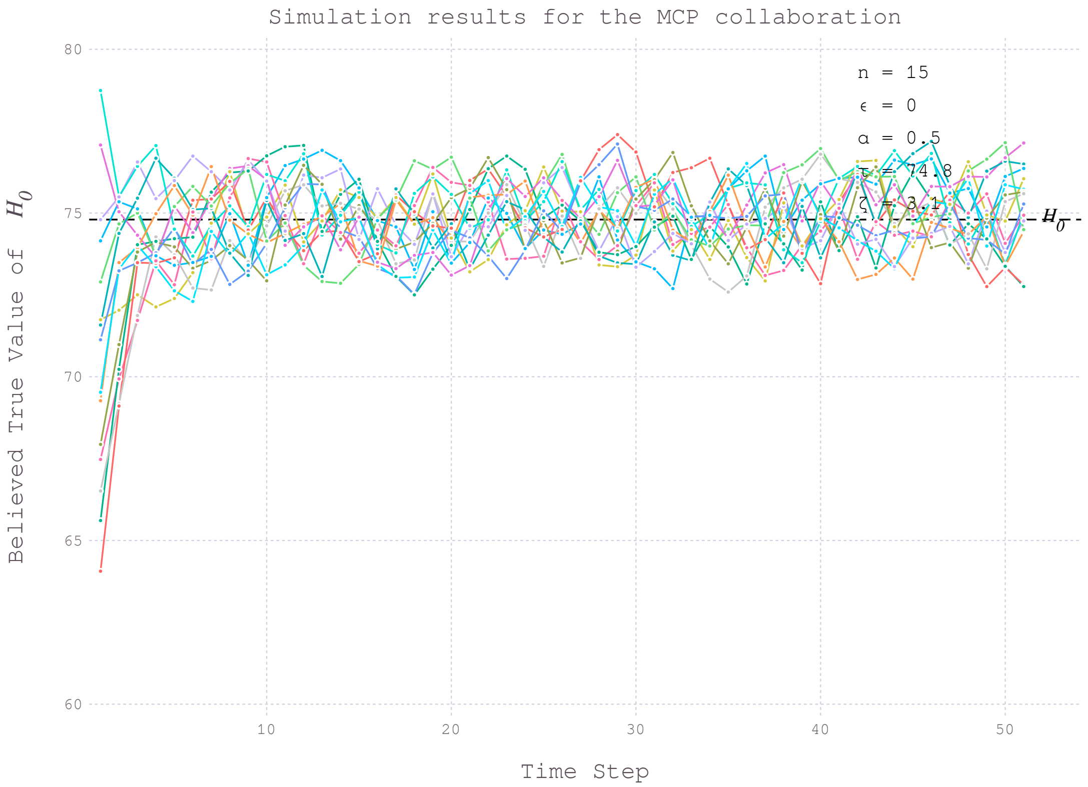

Figure 14: MCP collaboration with no difference-splitting. Note the very poor convergence here as well.

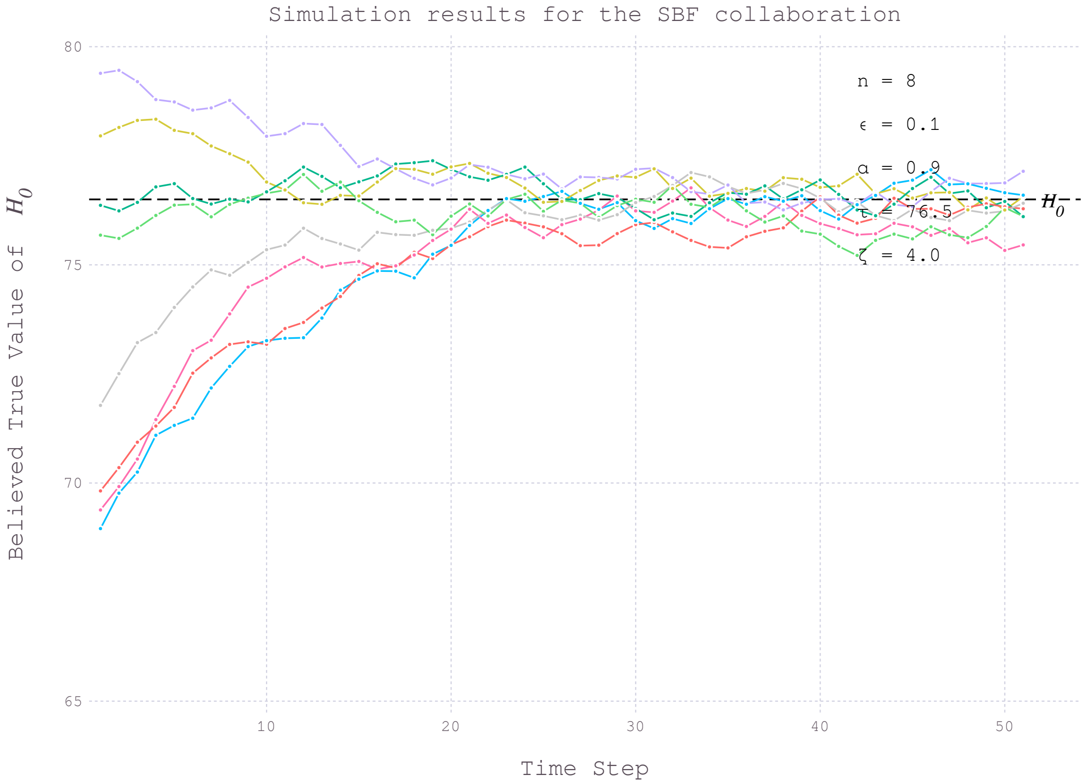

Figure 15: SBF collaboration with difference-splitting. Note the poor convergence due to the very small number of agents and the size of the noise level.

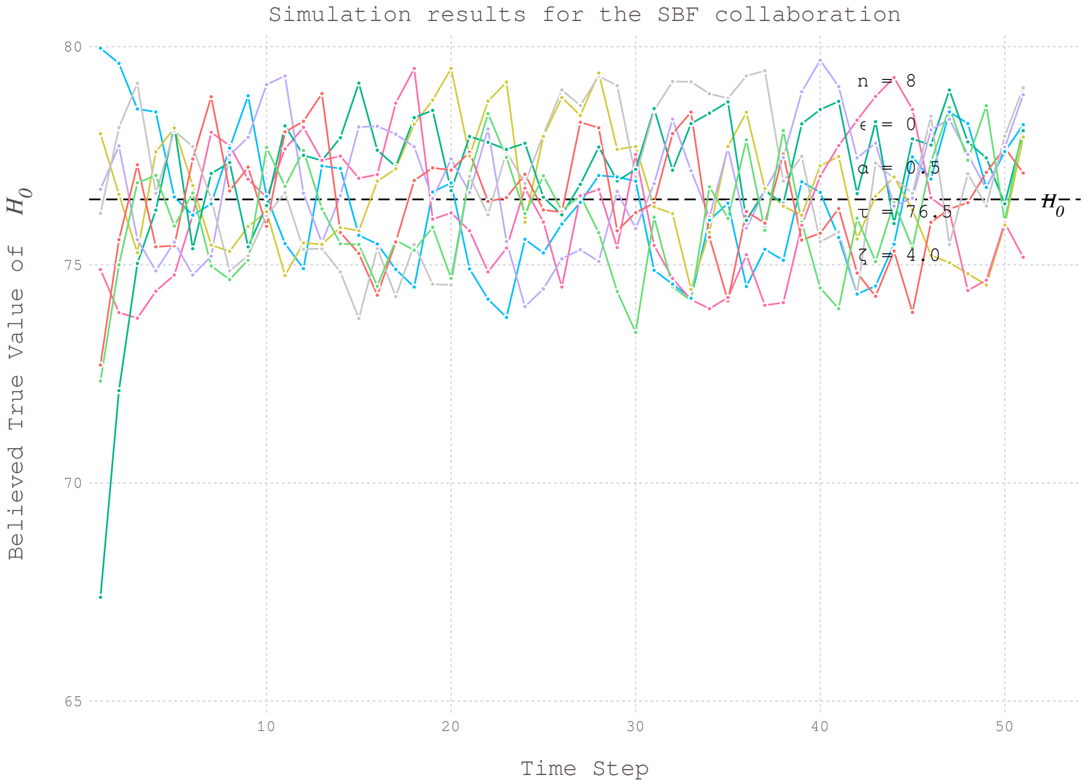

Figure 16: SBF collaboration with no difference-splitting, exhibiting the poorest convergence of the observation stage.

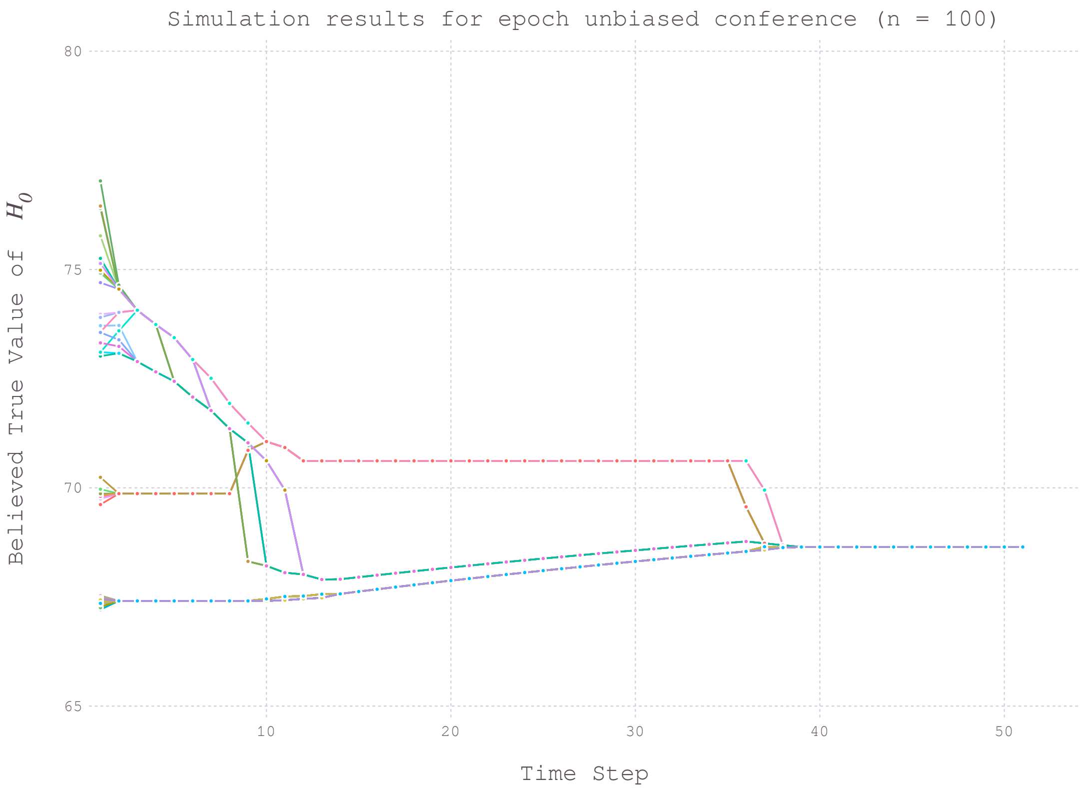

Figure 17: Unbiased conference converging on a single end-state. Note the complex dynamics of the late universe agents.

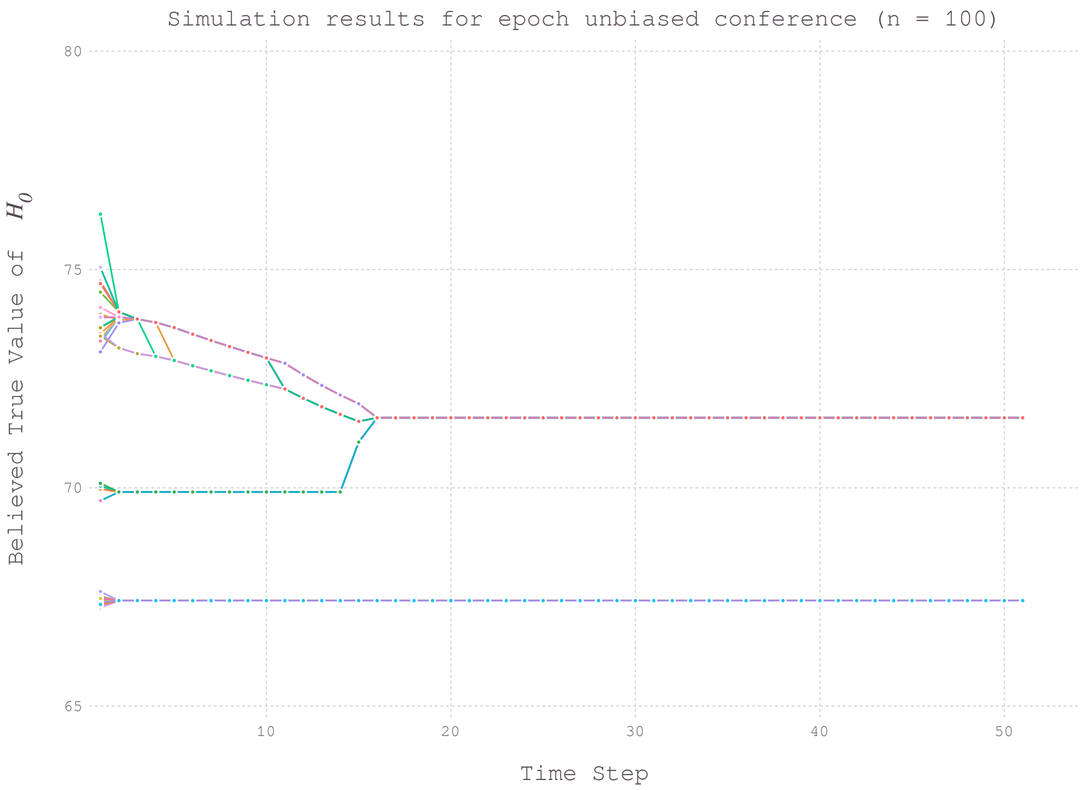

Figure 18: Unbiased conference resulting in two distinct end-states. Here the CCHP is absorbed into the late universe agents while the early universe agents do not interact with anyone else.

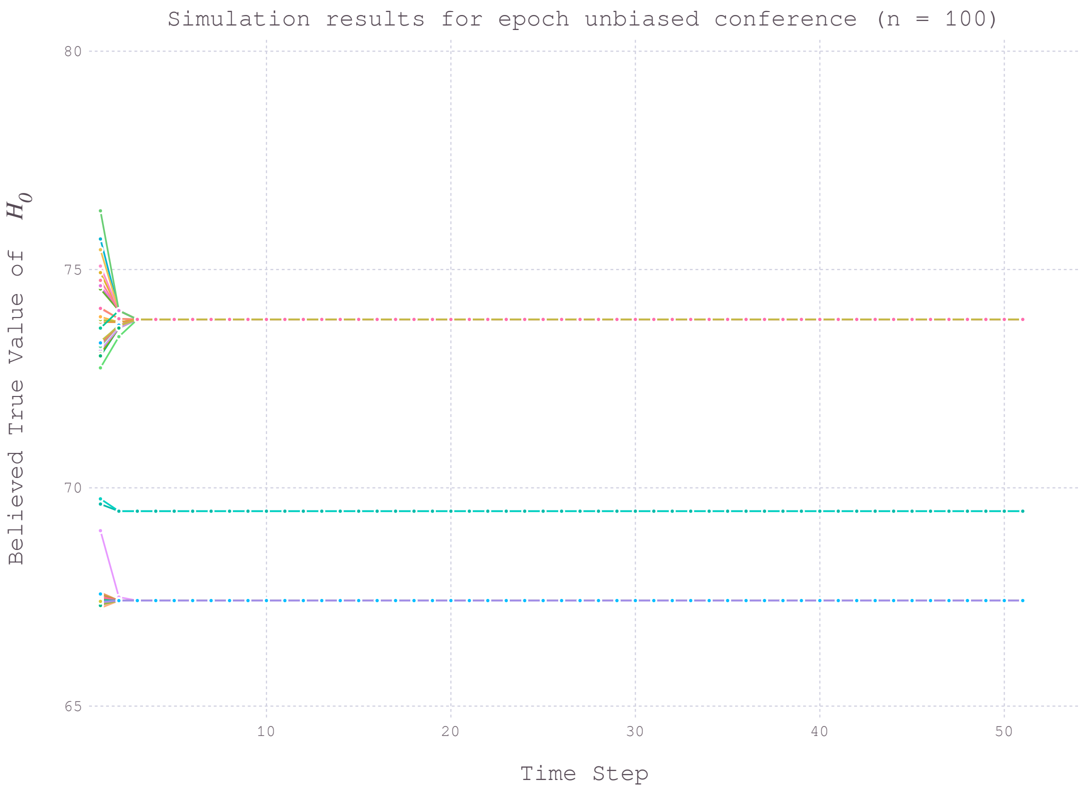

Figure 19: Unbiased conference resulting in three distinct end-states. Here neither the CCHP collaboration nor the early universe collaborations interact with late universe agents.

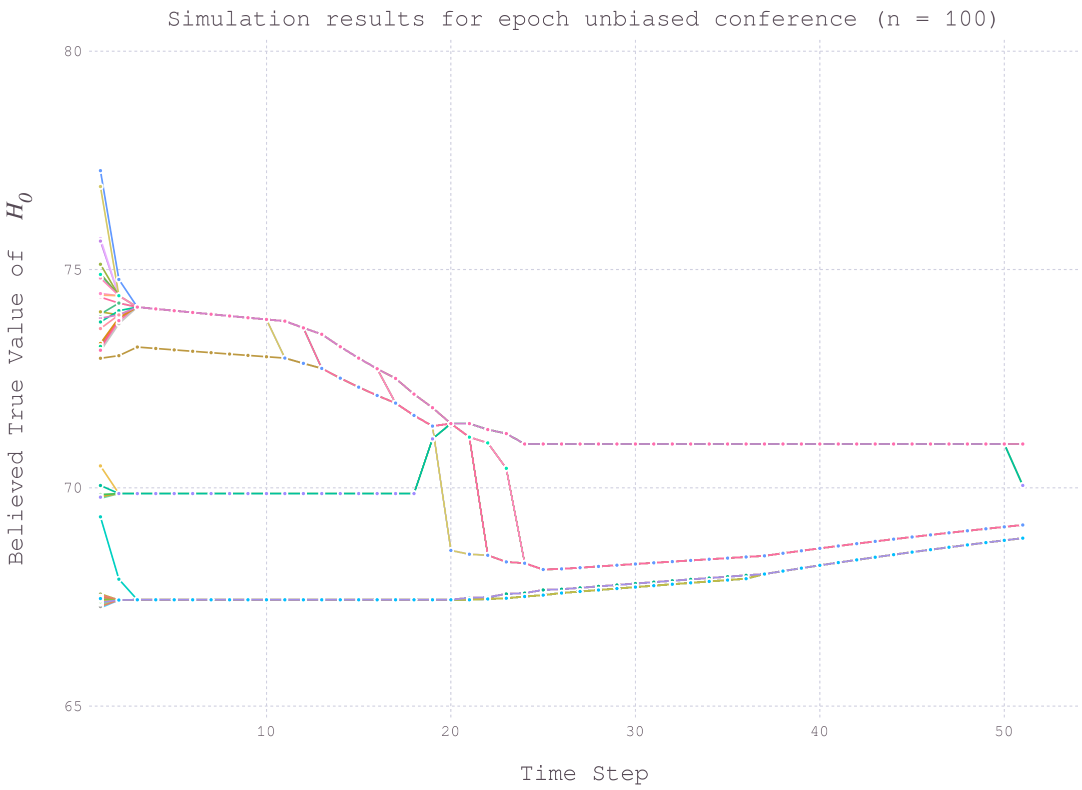

Figure 20: Unbiased conference resulting in four distinct end-states. Note the very complex dynamics implying the possibility of full convergence at a later time-step.

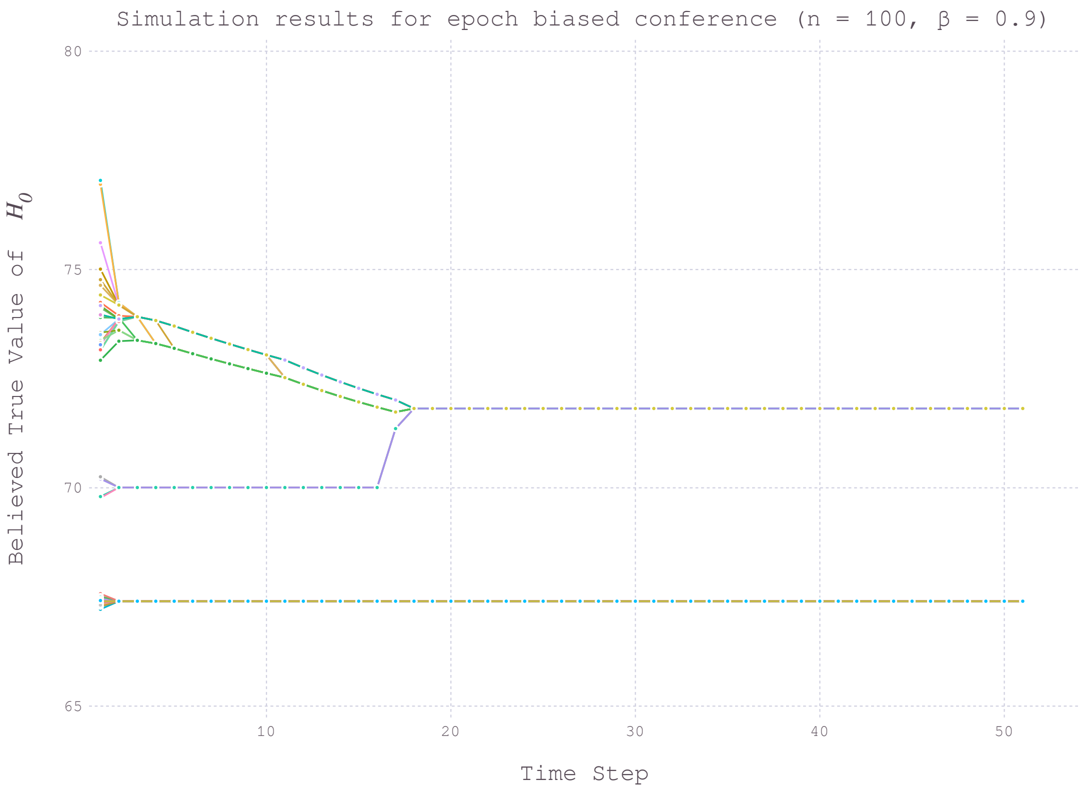

Figure 21: Biased conference resulting in two distinct end-states. Note the near identical structure to Figure 18.

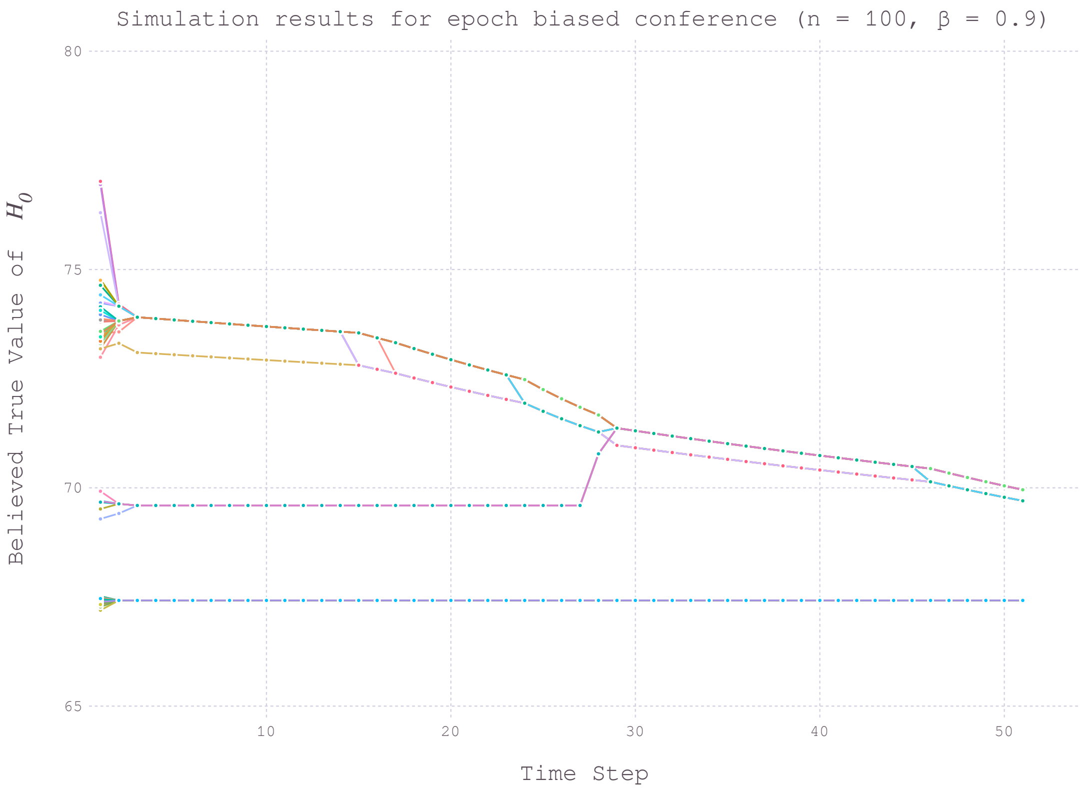

Figure 22: Biased conference resulting in three distinct end-states. Note how the late universe agents' interaction with the CCHP results in both getting attracted to the early universe agents.

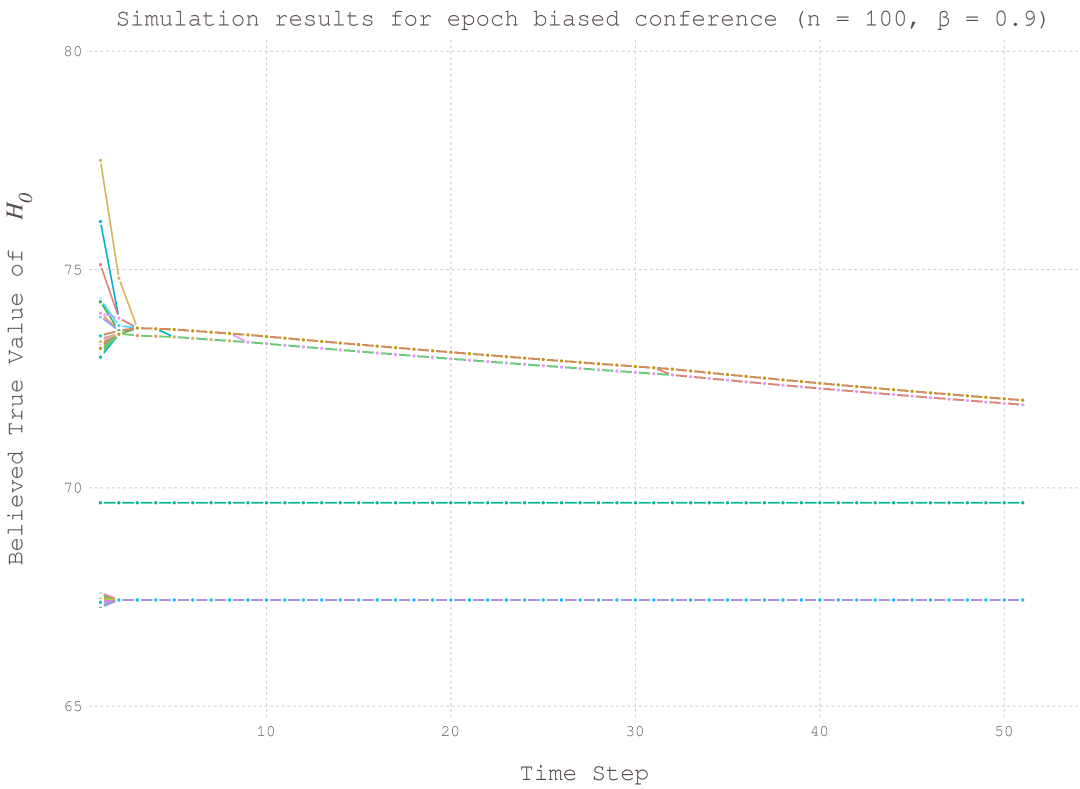

Figure 23: Biased conference resulting in four distinct end-states. Note that neither the CCHP nor the early universe collaborations interact with anyone outside their group.